Published in final edited form as:

J Clin Epidemiol. 2008 December ; 61(12): 1261–1270. doi:10.1016/j.jclinepi.2008.05.001.

Simple Counts of ADL Dependencies Do Not Adequately Reflect Older Adults' Preferences toward States of Functional Impairment

Tamara Sims, MA¹, Tyson H. Holmes, PhD³, Dena M. Bravata, MD, MS¹, Alan M. Garber, MD, PhD^{1,2}, Lorene M. Nelson, PhD⁴, and Mary K. Goldstein, MD, MS^{2,1}

¹ Center for Primary Care and Outcomes Research, Stanford University

² VA Palo Alto Health Care System

³ Division of Biostatistics, Health Research and Policy, Stanford University

⁴ Division of Epidemiology, Health Research and Policy, Stanford University

Abstract

Objective—To use unweighted counts of dependencies in Activities of Daily Living (ADLs) to assess the impact of functional impairment requires an assumption of equal preferences for each ADL dependency. To test this assumption, we analyzed standard gamble utilities of single and combination ADL dependencies among older adults. Study Design and Setting: Four hundred older adults used multimedia software (FLAIR1) to report standard gamble utilities for their current health and hypothetical health states of dependency in each of 7 ADLs and 8 of 30 combinations of ADL dependencies.

Results—Utilities for health states of multiple ADL dependencies were often greater than for states of single ADL dependencies. Dependence in eating, the ADL dependency with the lowest utility rating of the single ADL dependencies, ranked lower than 7 combination states. Similarly, some combination states with fewer ADL dependencies had lower utilities than those with more ADL dependencies. These findings were consistent across groups by gender, age, and education.

Conclusion—Our results suggest that the count of ADL dependencies does not adequately represent the utility for a health state. Cost-effectiveness analyses and other evaluations of programs that prevent or treat functional dependency should apply utility-weights rather than relying on simple ADL counts.

Keywords

Quality of Life; Older Adults; Activities of Daily Living; Standard Gamble Utilities; Functional Dependency; Health Preferences

What is new?

- Simple counts of the number of ADL dependencies, as often done for decision making about health care of older adults with functional limitations, are insufficient for assessing quality of life.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

- Health states of multiple dependencies ranked by simple counts do not correspond with rank by older adults' preference, or utility, ratings. Utilities for health states with the same count of ADL dependencies were often ordered differentially.
- Utility weights may be more appropriate than simple counts of the number of ADL dependencies for assessing the impact of functional dependency for policy and decision analyses.

1. Introduction

Between 1900 and 2004, the average life expectancy of a 65-year old living in the U.S. increased from 6 to nearly 18 years [1]. Current projections suggest that the percentage of older adults in the U.S. population will increase from 13% to 19% by 2050 [2]. Although the rate of disability has steadily declined in recent years [3], the absolute number of older adults with disabilities is expected to increase substantially. Evaluating the costs and benefits of interventions designed to enhance or maintain the functional status of older adults is important due to the escalating number of disabled older adults and because older adults rate physical function as the most important factor in maintaining optimal quality of life [4]. However, measuring functional status alone does not fully capture the effects of a health state or relevant treatments on quality of life [5].

Cost-effectiveness analyses (CEA) of health care interventions estimate the marginal cost per marginal health effect of the intervention compared to an alternative [6]. Health effects are typically measured in Quality Adjusted Life Years (QALYs). Health state utilities, also known as health preferences, may be applied as quality-weighting factors to estimate QALYs. Older adults' utilities for health states of functional dependency (i.e., dependency in at least one Activity of Daily Living (ADL) [7-8]) provide a method for appropriately valuing improvements in or maintenance of functional status when performing CEA of interventions that prolong or sustain independence among older adults. Yet, many researchers continue to rely on counts of ADL dependencies (without including an assessment of patient preferences) to measure functional impairment, which may undervalue their effectiveness. Using counts of ADL dependencies to assess functional status assumes that an individual does not value one ADL more than another, suggesting for example, that dependence in eating will have the same effect on one's quality of life as dependence in dressing. It also assumes that health states with fewer ADL dependencies are preferred to states with more ADL dependencies, regardless of the particular ADL dependencies.

Previous researchers have stressed the inadequacy of summarizing ADL counts as a measure of functional outcome [9-12] and some have argued for weighting ADL dependencies. For example, a panel of experts rated each ADL dependency's perceived importance according to the extent of human assistance each activity would require, which resulted in unequal weightings across the ADL dependencies (e.g., dependency in two lesser weighted ADLs are not necessarily equal to dependency in two heavily weighted ADLs) [10]. Many methods for applying weights to ADL dependencies either use this type of expert rating technique or arbitrary assignment [12]. However, expert ratings and arbitrary weighting may not capture the health preferences of older adults in the community [13-17]. For example, preferences for myocardial infarction reported in the Tufts CEA Registry [18], using a standard gamble utility scale from 0 (worst imaginable health state or death) to 1 (perfect health), showed clinician ratings to be 0.70 and community sample ratings to be 0.53. Moreover, Elstein and colleagues [13] found that clinician ratings of health states for prostate cancer corresponded with patient ratings only 50% of the time. An alternative approach to clinicians' weighting ADL dependencies is to elicit utilities for these health states directly from older adults. Unlike

arbitrary or indirect weights, such utilities provide a direct measurement of weights for evaluating treatments and programs designed for that population.

To evaluate the effect of multiple ADL dependencies on quality of life, we directly elicited older adults' standard gamble (SG) utilities. We have previously reported the utilities for ADL dependencies [19], which show that not all ADL dependencies are valued equally. Furthermore, preferences for health states of multiple dependencies (e.g., dependence in eating and in bathing concurrently) may not be directly calculable from preferences for health states of single dependencies, since some combinations may aggravate each other while other combinations may add little disability to the single component. In this study, we investigated the relationship between utility ratings for ADL dependencies and count of ADL dependencies in each health state. We also evaluated whether this relationship varies by patient characteristics including age, gender, and educational level.

2. Methods

Our methods of participant recruitment, utility assessment, and measurement of inconsistent and invariant responses have been described previously [19-20]. We briefly describe these in the sections that follow.

2.1. Recruitment of Older Adult Participants

We recruited 400 older adult members of the Kaiser Permanente Medical Care Program of Northern California, Santa Clara. We randomly selected individuals age 65 and older with recent medical appointments and notified them of the study in a letter mailed by Kaiser. The letter included a response card, which participants could return to accept or decline participation. We then contacted those who did not decline. We excluded individuals who were terminally ill, cognitively impaired, did not speak English, or who scored less than 24 on the Mini-Mental State Exam (MMSE) [21]. The study protocol received approval by Stanford University's Administrative Panels on Human Subjects in Medical Research and by Kaiser Permanente, Northern California Institutional Review Board Health Services Panel.

2.2. Multimedia Utility Assessment Software

To elicit standard gamble utilities for states of dependence in ADLs, we used Functional Life and Independence Rating (FLAIR1) multimedia software, a program specifically designed for computer-inexperienced older adults. This program combines text, graphics, and voice narration to present standardized health state descriptions and elicit SG utilities with minimal interviewer bias [20]. We initially pilot tested FLAIR1 (previously called "ADLIB") and reported these results elsewhere [22].

In the first section of the computerized assessment, we presented participants with descriptions of dependence (inability to perform activity without the help of another person) in each of the seven individual ADLs (bathing, continence, dressing, eating, toileting, transferring, and walking). Health state descriptions were comprised of detailed text and neutral photographs developed under the consult of health care professionals and older adults [22]. We then asked participants to report whether or not they were dependent in each ADL. Participants then rated the SG utility of their own current health. They next rated the SG utility of the health state of dependence in each of the seven ADLs presented one at a time. Research assistants underwent extensive training to assist participants if they experienced difficulty imagining health states or providing standard gamble ratings. We randomized the order of presentation of the seven single ADL dependency health states for each participant to minimize order effects.

After rating the health states of a single ADL dependency, participants rated a health state of dependence in all 7 ADLs in combination, followed by health states of dependence in 2 ADLs, 3 ADLs, 4 ADLs, 5 ADLs, and 6 ADLs. Early ADL scales suggested a hierarchical pattern in the development of functional limitations, establishing a general approach to the evaluation of health states of dependency [7-8]. However, subsequent research has demonstrated that patterns of multiple ADL dependencies do not necessarily correspond to this structure [11, 23-26]. In order to maximize ecological validity, we instead selected the 30 most commonly occurring ADL dependency combinations accounting for 98% of people at home with a functional dependency as reported by the National Long Term Care Survey [27]. To avoid participant fatigue, we only asked participants to rate a total of 15 out of the 37 possible health states: dependence in each of the seven single ADLs, dependence in the combination of all 7 ADLs, and a randomly selected seven health states of dependency from the remaining 29 ADL combinations. As depicted in Figure 1, these latter seven health states were randomly assigned to each participant by the FLAIR1 software under a modified fractional factorial design [28]. The FLAIR1 software presented participants with two opportunities to review and revise their ratings: after rating dependence in the single ADLs and the state of dependence in all 7 ADLs and again after rating the additional combination health states of dependency.

2.3. Additional Measures

Upon completion of the FLAIR1 program, participants were asked to provide demographic information (i.e., age, gender, education level, income level, and race/ethnicity). We also assessed their current physical and mental health using the Medical Outcomes Study Short Form-36 version 1 (SF-36v1) health status measure [29]. Additionally, we asked 154 participants to evaluate FLAIR1 using a scale ranging from 1 (disagree strongly) to 5 (agree strongly) for the following statements: “The questions asked in the computer program seemed clear or easily understood;” “The computer was easy to use;” “I felt the questions asked by the computer were confusing.”

2.4. Consistency Check

We randomly selected participants to receive consistency checks while providing their SG ratings. FLAIR1 evaluated responses for “potential inconsistencies” if the participant rated a combination of ADL dependencies higher than the rating for any of the single ADLs in that combination (e.g., a utility rating of 0.90 for dependence in bathing, dressing, and transferring, and a utility rating of 0.80 for dependence in bathing alone) [19]. Among patients randomized to receive the consistency checker, when they entered a “potentially inconsistent” response, a pop-up window informed the participant of the inconsistency but did not force a change in response. We use the term “potentially inconsistent” because some participants may have had a rational explanation [30] for valuing a single ADL dependency as less desirable than a combination set that included it. Here, we distinguish a “potential inconsistency” (i.e., giving an ADL dependency a lower utility than a health state in which that same ADL dependency is nested) from an inconsistency between the rank order of health states by number of ADL dependencies and the rank order of health states by utility ratings. The focus of this paper will be to examine the latter which may or may not encompass the former.

3. Data Analysis

3.1. Association between Counts and Utilities of ADL Dependencies

The purpose of this analysis was to evaluate whether the ranking of ADLs by counts was equivalent to the ranking of ADLs by utility ratings. First, we calculated the mean utility rating (and 95% confidence interval) for each health state. We then assessed whether there was high agreement between ordering of the health states according to the count of ADL dependencies and ordering of the health states according to mean utility ratings across participants. Due to

the complex sampling design (see Figure 1) in which each participant rated only 15 of the possible 30 health states, we applied a bootstrap method to perform this analysis [31]. Specifically, we drew 5000 samples of participants with replacement (i.e., a participant can be chosen several times within the same sample). For each re-sampling, we re-calculated the mean utility rating across all 400 participants for each health state. We then created two ordered sets of health states. The first set ordered health states by **count** of ADL dependencies and the second set ordered health states by mean utility **ratings** (e.g., the health state with the highest mean utility was ranked 1) for each of the 5000 re-samples. We then measured agreement between the two sets by comparing the number of ADL dependencies in the health states at corresponding rank positions. To do so, we calculated the absolute value of the difference between the number of ADLs in each of the two health states. We then summed these absolute differences across all comparisons as a measure of statistical agreement between the **count** and **ratings** sets. If the ordering of health states is the same when based on count of ADL dependencies or on mean utility ratings, the mean absolute difference summed across all rank positions would be zero. From the 5000 re-sample estimates of this agreement statistic, we calculated a 95% confidence interval on its mean using the percentile method [31] with correction for bias in the mean of the bootstrapped distribution.

3.2. Expected Ordering of Utilities

The purpose of the next analysis was to evaluate the extent to which participants provided lower utilities for states of fewer ADL dependencies than for states of more ADL dependencies across different combinations of ADL dependencies. These utility ratings may be potentially inconsistent, as when the ADL dependencies in the health states are nested, although not necessarily, as when the ADL dependencies in the health states are not nested. We first averaged each participant's actual utility ratings for all health states of 2 ADLs, for all health states of 3 ADLs, and so on (no averaging was needed for 7 ADL dependencies as this was only one state). We then calculated the proportion of participants who rated health states with a count of m ADL dependencies lower than or equal to health states with a count of $m+1$ ADL dependencies, where $m = 1, 2, 3, 4, 5$ and 6 . Given that each ADL dependency is weighted equally as assumed when using a simple counts approach, we would expect that m ADL dependencies would be rated higher on average than $m+1$ ADL dependencies. Note that because the set of fewer ADL limitations need not be a proper subset of the set of more ADL limitations, utility ratings could be greater for the state with more limitations without any inconsistency on the part of the participant.

For 1 versus 2 ADL dependencies, a participant was assigned a score of 0 if the average rating for health states of a single ADL dependency were rated higher than the average rating for health states of 2 ADL dependencies. If single ADL dependencies were rated equal to or lower on average than 2 ADL dependencies, the participant was assigned a score of 1. We repeated this scoring for each consecutive pair (e.g., assigning a score of 0 given that that average rating across 2 ADL dependencies was more than 3, 3 ADL dependencies was more than 4, 4 ADL dependencies was more than 5, 5 ADL dependencies was more than 6, and 6 ADL dependencies was more than 7), yielding six comparisons per participant, with each comparison scored on this 0/1 outcome. We performed logistic regression to estimate the proportion of participants (and its 95% confidence interval) that rated health states with the lesser ADL dependency count lower for each of the six comparisons. To account for the repeated measures structure, we used a nonlinear mixed effects model to regress the 0/1 outcome on five indicator variables, one for each comparison. We repeated this analysis by gender (men vs. women), education (<13 yrs. vs. ≥ 13 yrs.) and age group (65-74 yrs. vs. 75+ yrs.), in which we included a group variable as the sixth indicator.

3.3. Association between Utilities for Single ADL Dependencies and Combinations of ADL Dependencies

The purpose of this analysis was to evaluate the extent to which combination health states could be predicted by the single ADL dependency utilities. Specifically, we examined how accurately utility ratings for seven health states with a single ADL dependency predicted utility ratings for a health state consisting of all 7 ADL dependencies. For this analysis we assumed that the utilities for the single ADLs were independent. We calculated the product of the ratings of the seven single ADL dependencies for each participant. We assessed how well this product predicted the observed rating for the health state consisting of all 7 ADL dependencies.

3.4. Order Effects

The purpose of this analysis was to evaluate the extent to which participants might have been influenced by the order in which they were asked to provide utility ratings. The order of presentation of the seven single ADL dependency health states was randomized separately for each participant. We performed permutation testing [32] separately for each single ADL dependency health state to assess whether average utility ratings varied significantly among the six positions. A record of the order of presentation of the ADLs for each participant was not kept for FLAIR1. To assess for order effects in ratings of ADL dependencies, we used data from a related study in which older adults rated health states of ADL dependencies presented in randomized order [33].

4. Results

4.1. Participant characteristics

Participant ages ranged from 65 to 91 with a mean of 73 years (s.d. = 5.7); 56% (n = 222) were female; and 8% (n = 32) were non-white. Mean total years of education was 15 (s.d. = 2.7) and median category of annual income level was \$35,000-50,000. Ten percent of participants reported dependence in at least one of the seven ADLs. Mean SF-36v1 physical health component scores for our sample (44.75, s.d. = 9.90) were comparable to scores for U.S. adults 65 years and older (42.17, s.d. = 11.72) [29,34]. Mean SF-36v1 mental health component scores for this sample (54.08, s.d. = 8.2) and U.S. older adults (53.02, s.d. = 9.48) were also comparable [29,34].

Most participants reported having minimal or no computer experience. However, evaluations of the FLAIR1 program were favorable [20]: of 154 participant evaluations, 76% reported that the computer was easy to use. Moreover, participants appeared to have little to no difficulty providing standard gamble ratings for the health state descriptions, such that 77% thought the questions were clear and easily understood and only 15% felt the questions asked by the computer were confusing (research assistants were present to address confusion by providing additional clarification to these participants). The average length of time required to complete the FLAIR1 component of the interview was 36 minutes (s.d. = 15 minutes; median time = 33 minutes). Seventy-seven (19%) participants gave invariant responses—they were unwilling to accept any risk of death for any health state and reported a utility rating of 1.0 for all health states. One hundred and seventy-five (44%) participants gave a potentially inconsistent response for at least one health state and 96 (24%) of these participants reported a higher utility for dependence in all 7 ADLs than for a single ADL. We calculated the maximum magnitude of inconsistency as a participant's largest difference between an inconsistent rating of any single ADL and the rating of all 7 ADLs [19]. Across all participants, mean maximum magnitude of inconsistency was 0.17 (s.d. = 0.23).

4.2. Standard Gamble Utility Ratings

Standard gamble utilities for the entire sample ($n=400$) have been previously reported [19]. On the conventional utility scale in which 1 = perfect health and 0 = death or the worst imaginable health state, the mean utility rating for current health was 0.90. Because utility ratings are bounded by a 0 to 1 scale, we used nonlinear regression analysis [35] to compute a 95% confidence interval for the mean rating of current health. This confidence interval's lower bound was 0.87 and upper bound was 0.92, which is clearly below 1, as we anticipated for a sample of older adults. Means and standard deviations of utility ratings for the health states of single ADL dependencies are summarized in Table 1.

In Table 2, health states of combinations of ADL dependencies are ordered from highest to lowest mean utility. As the table demonstrates, the mean utility ratings for combination states with fewer ADL dependencies are often ranked lower than combination states with more ADL dependencies. For example, dependence in eating alone, the single ADL health state with the lowest utility rating (0.85), had a mean utility that ranked lower than six of the combination states. By visual inspection of Table 2, no single ADL dependency is included in all combinations of dependency. Thus, we cannot infer that the ratings of combinations of ADL dependencies are being driven by a particular single ADL dependency.

4.3. Association between Counts and Utilities of ADL Dependencies

We used our bootstrap method to test whether the ordering of health states based on mean utility ratings vs. counts of ADL dependencies differed significantly. The mean absolute difference in number of ADLs between health states ordered by count and those ordered by mean utility ratings is 1.08 (i.e., roughly one ADL). The 95% confidence interval on this mean absolute difference (lower = 0.82, upper = 1.31) is bounded well away from zero, indicating statistically significant disagreement ($p < 0.05$) in the ordering of health states based on count of ADL dependencies compared to the ordering of health states based on mean utility ratings.

4.4. Expected Ordering of Utilities

Table 3 presents the percentage of participants who rated m ADL dependencies lower than or equal to $m + 1$ ADL dependencies for $m = 1, 2, 3, 4, 5$ and 6. We found that 63% (95%CI: 57%-69%) of participants rated health states consisting of a single dependency lower than or equal to health states of two dependencies (Table 3). We also found that the percentage of participants rating health states of fewer dependencies lower than or equal to health states of more dependencies are also high for 2 dependencies vs. 3, 3 dependencies, vs. 4, 4 dependencies vs. 5 and 5 dependencies vs. 6 (Table 3), all contrary to expected rating patterns based on ADL counts alone. Interestingly, 70% (95%CI: 64%-77%) of the participants rated health states consisting of 6 ADL dependencies lower than or equal to health states of all 7 ADL dependencies.

We evaluated the order of mean utilities for men vs. women, for participants under/over 75 years, and for participants with more/less than 13 years of education. For all subgroups, mean utilities for health states with fewer dependencies were occasionally ranked lower than health states with more dependencies. For instance, the combination health state of dependence in toileting and continence was ranked lower on average than dependence in dressing, bathing, and toileting, for all subgroups by age, gender and education.

To assess whether the percentage of participants rating health states with fewer dependencies lower than health states with more dependencies was statistically significant, we performed additional nonlinear mixed-effects regression analyses as described above by age, gender, and education level. Analyses revealed similar response patterns across groups (see Tables 4-6). We did not find a statistically significant difference between age groups or between education

groups in the percentage of participants rating health states with fewer dependencies lower on average than health states with more dependencies; however, this percentage was significantly higher for women than men ($p = 0.039$).

4.5. Association between Utilities for Single ADL Dependencies and Combinations of ADL Dependencies

There was no difference between the predicted rating of a combination of 7 ADL dependencies (i.e., the product of the ratings for each of the seven single ADL dependencies) and the observed rating for 22% of the participants, indicating perfect agreement. Overall, most differences were near 0 (see Figure 2), illustrating high concordance between the predicted and observed utility rating for the combination health state. The distribution was somewhat skewed, with more positive differences than negative differences, meaning that the product of the observed ratings of the seven single ADL dependencies tends to predict a lower utility for a combination of 7 ADL dependencies than the observed rating. The St. Laurent's coefficient [36] between the predicted and observed ratings also demonstrated high agreement at 0.74, where 1 is perfect agreement.

4.6. Removing Participants with Invariant Responses from the Analyses

Nineteen percent of participants gave invariant responses, which may account for a substantial proportion of participants rating health states with fewer ADL dependencies lower than *or equal to* health states with more ADL dependencies. When removing invariant participants from the analysis, the proportion of participants rating states with fewer ADL dependencies lower than or equal to more ADL dependencies did decrease for most comparisons (see Figure 3), however the proportions remained significantly high, $p < 0.05$. Of the 29 participants with invariant responses, 34% reported that the questions were confusing, whereas only 10% who did not give invariant responses ($n=125$) found the questions confusing. We also assessed whether confusion with the program accounts for rating health states with fewer dependencies lower than or equal to more dependencies. To do so, we performed nonlinear mixed-effects regression analysis as described in Section 3.1, except we added “confusion” as the sixth indicator variable. Although an odds ratio of 0.45 indicated that confusion increases the likelihood of rating fewer dependencies lower or equal to more dependencies, the difference was not statistically significant, (95% CI, 0.20 to 1.01).

4.7. Order Effects

Finally, we assessed whether mean utility ratings for single ADL dependencies varied as a function of presentation order. We found no significant effect of the order for five of the six health states ($p > 0.14$), with dependence in toileting being the possible exception (0.04).

5. Discussion

Our study suggests that a simple count of ADL dependencies is not a good proxy for the quality of life associated with a health state. The rank ordering of ADL dependencies by count does not match the rank order by utility rating. Using the standard gamble method, we were able to determine that a substantial proportion of participants rated health states with fewer dependencies equal to or even lower than more dependencies. These findings indicate that older adults differentially value independence in ADLs, placing greater utility on some more than others. They also differentially value independence in combinations of ADLs, placing greater utility on some combinations than others with the same count of ADLs. Moreover, dependence in 2 or 3 ADLs can lead to a greater loss of utility than dependence in 4 or 5. Even dependence in a single ADL can result in a lower utility rating than dependence in some combinations of ADLs. Therefore, older adults' health preferences for multiple dependencies cannot be estimated simply from calculating the total count of ADL dependencies. These

findings were consistent across gender, age, and education. Importantly, we did find that utility for a health state with 7 ADL dependencies can be accurately estimated from utilities for single ADL dependencies, suggesting that future work may not need to directly assess all possible combinations of ADL dependencies. Our data suggest that while counts of ADL dependencies are insufficient, SG utilities may be a more appropriate and comprehensive method for measuring the impact of functional limitations on quality of life.

Confusion with the questions was moderately associated with invariance, or giving a utility of 1.0 to all health states. However, confusion did not significantly predict rating fewer dependencies lower than or equal to more dependencies. Furthermore, when removing invariant participants, a substantial proportion rating fewer dependencies lower than or equal to more dependencies remained. Thus, even among those willing to take some risk, relying on simple counts of ADL dependencies is still inappropriate.

Disagreement in ranking of health states between ADL counts and utilities may be partly due to potentially inconsistent responses described earlier. Although we do not examine inconsistent responses specifically here, research suggests that “inconsistent” responses may reflect true preferences [30]. In fact, in a current extension of this study, several participants remarked that the possibility of receiving professional care would increase their utility for several ADL dependencies compared to 1 or 2 dependencies. Such findings underscore the value of the SG method as it is able to capture multifaceted responses that other scaling methods cannot. To better understand such patterns of responding, we are currently evaluating modifications of the FLAIR program (FLAIR2) which include an additional method of valuation for comparison to the SG (i.e., willingness to pay) [33].

Although the education and income level of our sample was higher than the national average [1] and was predominantly Caucasian, their physical and mental health status was representative of U.S. older adults [25]. Previous research has shown that current health has a significant impact on health state valuations [37], which outweigh the effects of participant characteristics such as race and income on SG ratings of hypothetical health states [38]. Because the health status of our sample is characteristic of U.S. older adults, our findings should generalize to the older adult population.

Mean utilities for each health state across individuals ranged from 0.76 to 0.90 while standard deviations ranged from 0.13 to 0.34, suggesting substantial variability within utilities. We recognize that an increased sample size may reduce such variability. However, it may be that standard gamble utility ratings for ADL dependencies are inherently variable, indicating the sensitivity of the SG method. In looking at the difference between utilities, even the slightest difference may have a large effect on decision analysis or CEA results, which are highly context-specific and vary in the degree of sensitivity to the utility depending on the treatment design, costs, and outcomes. Utilities from several studies are posted on the Tufts University CEA Registry website [18]. Focusing on utilities elicited by the standard gamble method, one sees by inspection that a difference of 0.08 in utility is meaningful (e.g., mild Amyotrophic Lateral Sclerosis [ALS], or Lou Gehrig's disease, is reported as having a utility of 0.79 and severe ALS has a utility of 0.71), suggesting that a utility difference of less than 0.10 is clinically significant. The substantial impact of slight variation also highlights the importance of directly assessing the target population for which a treatment is designed.

6. Conclusion

Overall, our findings indicate that an unweighted count of ADL dependencies may be an insufficient measure of quality of life. Researchers and health resource allocation agencies rely on simple counts of ADL dependencies for formal evaluation of programs designed to prevent

or treat functional dependency in the aging population. However, we find that more general approaches to utility measurement yield different rankings of health states than do ADL counts and that health states cannot be valued in terms of simple counts of ADL limitations.

Furthermore, summary measurements of functional status (e.g., IADL counts, number of medical conditions) cannot take into account individual values and therefore do not provide a comprehensive assessment of quality of life. Preference-based utilities capture a more complex picture of the value older adults place on functional independence. Population status assessments and program evaluations should apply appropriate utility weights for the relevant combinations of ADL limitations for a more accurate representation of a particular health state's value.

Acknowledgements

This work was supported by National Institute on Aging grants R01 AG15110 (PI: Goldstein), P30 AG 17253 (Garber, PI), and P30 AG24957-01 (Garber, PI), as well as the Department of Veterans Affairs. Dr. Goldstein's work was supported in part by VA HSR&D RCD-96-301 and ARCD-96301. The authors wish to thank the Northern California Kaiser Permanente Medical Care Program (KPMCP) members for participating in this study and the FLAIR staff at the Division of Research, KPMCP. The authors also thank Sheryl Davies for FLAIR1 project management, Pamela Mahlow for data management, and David Miller for FLAIR1 programming. An early version of the manuscript was presented at the Bay Area Health Care Quality and Outcomes Research Conference; the authors thank Sara Knight, PhD, for helpful commentary. The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs.

References

1. Federal Interagency Forum on Aging Related Statistics. Older Americans 2004: Key Indicators of Well-Being. Washington D.C.: U.S. Government Printing Office; Nov. 2004
2. Dawson, R.; Hendershott, G.; Fulton, J. Aging in the Eighties: functional limitations of individuals age 65 years and over. Washington: Ntl Ctr Hlth Stats; 1987.
3. Manton KG, Gu X, Lamb VL. Change in chronic disability from 1982 to 2004/2005 as measured by long-term care changes in function and health in the U.S. elderly population. PNAS 2006;103(48): 18374–18379. [PubMed: 17101963]
4. Sutherland HJ, Lockwood GA, Boyd NF. Ratings of the importance of quality of life variables. J Clin Epidemiol 1990;43:661–666. [PubMed: 2370573]
5. King JT, Tsevat J, Roberts MS. Preference-Based Quality of Life in Patients with Cerebral Aneurysms. Stroke 2005;36:303–309. [PubMed: 15653579]
6. Garber, AM.; Weinstein, MC.; Torrance, GW.; Kamlet, MS. Theoretical Foundations of Cost-Effectiveness Analysis. In: Gold, MR.; Siegel, JE.; Russell, LB.; Weinstein, MC., editors. Cost-Effectiveness in Health and Medicine. New York: Oxford University Press; 1996. p. 25-53.
7. Katz S, Akpom CA. 12. Index of ADL. Med Care 1976;14(5 Suppl):116–118. [PubMed: 132585]
8. Katz S, Ford AB, Moskowitz RW, Jackson BA, Jaffee MW. Studies of illness in the aged. The index of ADL: a standardized measure of biological and psychosocial function. J Am Med Assoc 1963;185 (12):914–919.
9. Thomas VS, Rockwood K, McDowell I. Multidimensionality in Instrumental and Basic Activities of Daily Living. J Clin Epidemiol 1998;51(4):315–321. [PubMed: 9539888]
10. Finch M, Kane RL, Philp I. Developing a new metric for ADLs. J Am Geriatr Soc 1995;43(8):877–884. [PubMed: 7636095]
11. Travis SS, McAuley WJ. Simple counts of the number of basic ADL dependencies for long-term care research and practice. Health Serv Res 1990;25(2):349–360. [PubMed: 2113045]
12. Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. Maryland State Med J 1965;14:61–65.
13. Elstein AS, Chapman GB, Knight SJ. Patients' Values and Clinical Substituted Judgments: The Case of Localized Prostate Cancer. Health Psych 2005;24(4 Suppl):S85–92.

14. Chen Q, Kane RL. Effects of using consumer and expert ratings of an activities of daily living scale on predicting functional outcomes of postacute care. *J Clin Epidemiol* 2001;54(4):334–342. [PubMed: 11297883]
15. Kane RL, Rockwood T, Finch M, Philp I. Consumer and professional ratings of the importance of functional status components. *Health Care Financing Rev* 1998;19(2):11–22.
16. Philp I, Armstrong GK, Coyle GG, Chadwick I, Machado ABC. A better way to measure disability in older people. *Age and Ageing* 1998;27:519–522. [PubMed: 9884012]
17. Kane RL, Rockwood T, Philp I, Finch M. Differences in Valuation of Functional Status Components among Consumers and Professionals in Europe and the United States. *J Clin Epidemiol* 1998;51(8): 657–666. [PubMed: 9743314]
18. Tufts-New England Medical Center Institute for Clinical Research and Health Policy Studies. The Cost-Effectiveness Analysis Registry [database on the Internet]. C1998-2001 – [cited 2008 Feb 1]. Available from <http://www.tufts-nemc.org/cearegistry/data/phaseIIpreferenceweights.pdf>
19. Bravata DM, Nelson LM, Garber AM, Goldstein MK. Invariance and Inconsistency in Utility Ratings. *Med Decis Making* 2005;25:158–167. [PubMed: 15800300]
20. Goldstein MK, Miller DE, Davies S, Garber AM. Quality of Life Assessment Software for Computer-Inexperienced Older Adults: Multimedia Utility Elicitation for Activities of Daily Living. *Proc AMIA Symp* 2002:295–299. [PubMed: 12463834]
21. Folstein MF, Folstein SE, McHugh PR. “Mini-Mental State:” a practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975;12:189–198. [PubMed: 1202204]
22. Goldstein MK, Michelson D, Clarke AE, Lenert LA. A multimedia preference-assessment tool for functional outcomes. *Proc Annu Symp Comput Appl Med Care* 1993:844–848. [PubMed: 8130598]
23. Thomas VS, Rockwood K, McDowell I. Multidimensionality in Instrumental and Basic Activities of Daily Living. *J Clin Epidemiol* 1998;51(4):315–321. [PubMed: 9539888]
24. Kempen GJIM, Myers AM, Powell LE. Hierarchical structure in ADL and IADL: Analytical assumptions and applications for clinicians and researchers. *J Clin Epidemiol* 1995;48(11):1299–1305. [PubMed: 7490592]
25. Lazaridis EN, Rudberg MA, Furner SE, Cassel CK. Do Activities of Daily Living Have a Hierarchical Structure? An Analysis Using the Longitudinal Study of Aging. *J Gerontol* 1994;49(2):M47–M51. [PubMed: 8126352]
26. Siu A, Reuben D, Hays R. Hierarchical Measures of Physical Function in Ambulatory Geriatrics. *J Am Geriatr Soc* 1990;38:1113–1119. [PubMed: 2229865]
27. NLTCS. Limitations in Activities of Daily Living Among the Elderly: Data Analyses from the 1989 National Long-Term Care Survey. Washington, D.C.: National Aging Information Center, Administration on Aging, U.S. Department of Health and Human Services; May. 1996
28. Kotz, S.; Johnson, NL.; Read, CB. *Encyclopedia of Statistical Sciences*. 3. New York, NY: John Wiley & Sons, Inc.; 1983.
29. Ware, JE.; Kosinski, M.; Gandek, B. SF-36® Health Survey: Manual and Interpretation Guide. Lincoln, RI: Quality Metric Incorporated; 1993 2000.
30. Steward D, Davis M. Qualification of Discordant Responses in Utility Assessment. *Proc AMIA Symp* 1998:195–199. [PubMed: 9929209]
31. Efron, B.; Tibshirani, RJ. *An Introduction to the Bootstrap*. New York, NY: Chapman & Hall; 1993.
32. Good, P. *Permutation, Parametric, and Bootstrap Tests of Hypotheses: A Practical Guide to Resampling Methods for Testing Hypotheses*. Third. New York, NY: Springer-Verlag; 1994.
33. Sims TL, Garber AM, Miller DE, Mahlow PT, Bravata DM, Goldstein MK. Multimedia Quality of Life Assessment: Advances with FLAIR. *Proc AMIA Symp* 2005:694–698. [PubMed: 16779129]
34. Ware, JE.; Kosinski, M.; Dewey, JE. *How to Score Version 2 of the SF-36® Health Survey*. Lincoln, RI: Quality Metric Incorporated; 2000.
35. Neter, J.; Kutner, MH.; Nachtsheim, CJ.; Wasserman, W. *Applied Linear Statistical Models*. Fourth. Boston, MA: WCB McGraw-Hill; 1996.
36. St Laurent RT. Evaluating Agreement with a Gold Standard in Method Comparison Studies. *Biometrics* 1998;54:537–545. [PubMed: 9629642]

37. Dolan P. The Effect of Experience of Illness on Health State Valuations. *J Clin Epidemiol* 1996;49 (5):551–564. [PubMed: 8636729]
38. King JT, Tsevat J, Roberts MS. Positive Association between Current Health and Health Values for Hypothetical Disease. *Med Decis Making* 2004;24:367–378. [PubMed: 15271275]

Figure 1.

Example of ADL combination health states rated as assigned using a randomized fractional factorial design. This figure shows several examples of health states with combinations of more than one ADL dependency. The first line shows the health states in FLAIR1 that have a combination of 2 ADL dependencies, the next line shows the health states with a combination of 3 ADL dependencies, and so forth. For each participant using FLAIR1 to provide utility ratings, the computer program randomly selects 2 of the 2 ADL dependency combinations out of a possible 9, 2 of the 3 ADL dependency combinations out of a possible 7, 1 of the 4 ADLs out of a possible 6, 1 of the 5 ADLs out of a possible 5, and 1 of the two possible 6 ADL dependency combinations. For example, a participant may be presented with Transferring and Dressing for one of the 2 ADL dependency combinations and Toileting and Continence for the other.

Figure 2.

Percent of agreement between predicted rating of a health state with all 7 ADL dependencies (i.e., product of single ADL dependency ratings) and the observed rating. Value of zero indicates perfect agreement between predicted and observed; negative values indicate that participants gave a lower utility for the combination health state than the product of each of the seven single ADL dependencies; positive values indicate that participants gave a higher utility for the combination health state than the product of each of the seven single ADL dependencies.

Figure 3.

Comparison of samples including invariants and excluding invariants for proportion of participants with averaged utilities for combination health states with fewer ADL dependencies rated lower than those with more ADL dependencies.

Table 1

Utilities for Single ADL Dependencies (n=400)

ADL Dependency	Mean	SD
Dressing	0.896	0.219
Bathing	0.884	0.230
Continence	0.868	0.240
Transferring	0.868	0.244
Toileting	0.860	0.248
Walking	0.859	0.244
Eating	0.852	0.259

Table 2
Mean Utilities for Combination ADL Dependencies

ADL Dependency	# of ADLs	Mean	95% Confidence Interval (CI)	
			Lower	Upper
Walking and Transferring	2	0.892	0.851	0.934
Continence and Dressing	2	0.889	0.850	0.928
Dressing, Bathing, Toileting *	3	0.878	0.838	0.918
Bathing, Toileting, Continence *	3	0.875	0.835	0.915
Bathing and Continence	2	0.857	0.813	0.901
Transferring and Dressing	2	0.856	0.813	0.899
Walking, Dressing, Bathing, Toileting *	4	0.850	0.790	0.910
Transferring, Dressing, Bathing, Continence *	4	0.848	0.798	0.897
Walking, Dressing, Bathing *	3	0.841	0.790	0.892
Walking and Continence *	2	0.837	0.788	0.886
Walking, Dressing, Bathing, Continence *	4	0.834	0.770	0.897
Transferring, Walking, Dressing, Bathing, Continence	5	0.833	0.773	0.893
Bathing and Dressing *	2	0.829	0.764	0.894
Transferring, Walking, Dressing, Bathing, Toileting, Continence	6	0.828	0.782	0.874
Toileting and Continence *	2	0.828	0.763	0.892
Transferring and Bathing *	2	0.825	0.767	0.884
Dressing, Bathing, Toileting, Continence	4	0.824	0.752	0.896
Transferring, Walking, Bathing, Toileting, Continence	5	0.818	0.756	0.880
Transferring, Dressing, Toileting, Bathing, Continence	5	0.816	0.755	0.878
Walking, Bathing, Continence *	3	0.815	0.769	0.861
Transferring, Walking, Dressing *	3	0.815	0.759	0.871
Bathing, Continence, Dressing *	3	0.808	0.754	0.862
Transferring, Walking, Bathing *	3	0.807	0.754	0.860
Transferring, Walking, Dressing, Bathing, Toileting	5	0.790	0.724	0.856
Transferring, Walking, Dressing, Bathing *	4	0.788	0.712	0.863
Walking and Bathing *	2	0.787	0.722	0.852
Eating, Transferring, Walking, Dressing, Bathing, Toileting *	6	0.779	0.731	0.826
Transferring, Dressing, Bathing, Toileting *	4	0.758	0.676	0.840
Walking, Dressing, Bathing, Toileting, Continence *	5	0.758	0.691	0.825
All 7 ADLs	7	0.758	0.726	0.789

* Out of order as would be expected by count of ADL dependencies

Table 3

Proportion of participants with averaged utilities for combination health states with fewer ADL dependencies rated lower than or equal to those with more ADL dependencies.

Lower Level Combinations vs. Higher Level Combinations	Proportion Estimate	95% CI	
		Lower	Upper
Proportion Single Ratings $\leq$ 2 ADL Ratings	0.627	0.565	0.688
Proportion of 2 ADL Ratings $\leq$ 3 ADL Ratings	0.766	0.715	0.818
Proportion of 3 ADL Ratings $\leq$ 4 ADL Ratings	0.760	0.708	0.812
Proportion of 4 ADL Ratings $\leq$ 5 ADL Ratings	0.811	0.765	0.857
Proportion of 5 ADL Ratings $\leq$ 6 ADL Ratings	0.796	0.742	0.850
Proportion of 6 ADL Ratings $\leq$ All 7 ADLs Combined	0.700	0.635	0.765

Table 4

Proportion of older old versus younger old participants with averaged utilities for combination health states with fewer ADL dependencies rated lower than or equal to those with more ADL dependencies.

Age Group	Lower Level Combinations vs. Higher Level Combinations	Proportion Estimate	95% CI	
			Lower	Upper
65-74 yrs	Proportion Single Ratings $\leq$ 2 ADL Ratings	0.642	0.562	0.721
	Proportion of 2 ADL Ratings $\leq$ 3 ADL Ratings	0.714	0.641	0.788
	Proportion of 3 ADL Ratings $\leq$ 4 ADL Ratings	0.751	0.682	0.820
	Proportion of 4 ADL Ratings $\leq$ 5 ADL Ratings	0.784	0.720	0.848
	Proportion of 5 ADL Ratings $\leq$ 6 ADL Ratings	0.785	0.711	0.859
	Proportion of 6 ADL Ratings $\leq$ All 7 ADLs	0.715	0.630	0.800
	Combined			
75+ yrs	Proportion Single Ratings $\leq$ 2 ADL Ratings	0.596	0.497	0.696
	Proportion of 2 ADL Ratings $\leq$ 3 ADL Ratings	0.844	0.778	0.909
	Proportion of 3 ADL Ratings $\leq$ 4 ADL Ratings	0.775	0.696	0.855
	Proportion of 4 ADL Ratings $\leq$ 5 ADL Ratings	0.855	0.792	0.918
	Proportion of 5 ADL Ratings $\leq$ 6 ADL Ratings	0.815	0.738	0.893
	Proportion of 6 ADL Ratings $\leq$ All 7 ADLs	0.677	0.576	0.778
	Combined			

Table 5

Proportion of male versus female participants with averaged utilities for combination health states with fewer ADL dependencies rated lower than or equal to those with more ADL dependencies.

Gender	Lower Level Combinations vs. Higher Level Combinations	Proportion Estimate	95% CI	
			Lower	Upper
Male	Proportion Single Ratings $\leq$ 2 ADL Ratings	0.608	0.517	0.699
	Proportion of 2 ADL Ratings $\leq$ 3 ADL Ratings	0.712	0.629	0.794
	Proportion of 3 ADL Ratings $\leq$ 4 ADL Ratings	0.671	0.585	0.758
	Proportion of 4 ADL Ratings $\leq$ 5 ADL Ratings	0.769	0.695	0.843
	Proportion of 5 ADL Ratings $\leq$ 6 ADL Ratings	0.769	0.686	0.853
	Proportion of 6 ADL Ratings $\leq$ All 7 ADLs Combined	0.691	0.596	0.786
Female	Proportion Single Ratings $\leq$ 2 ADL Ratings	0.642	0.558	0.727
	Proportion of 2 ADL Ratings $\leq$ 3 ADL Ratings	0.810	0.746	0.873
	Proportion of 3 ADL Ratings $\leq$ 4 ADL Ratings	0.827	0.767	0.888
	Proportion of 4 ADL Ratings $\leq$ 5 ADL Ratings	0.845	0.788	0.901
	Proportion of 5 ADL Ratings $\leq$ 6 ADL Ratings	0.818	0.748	0.888
	Proportion of 6 ADL Ratings $\leq$ All 7 ADLs Combined	0.708	0.619	0.797

Table 6

Proportion of less educated versus more educated participants with averaged utilities for combination health states with fewer ADL dependencies rated lower than or equal to those with more ADL dependencies.

Education Level	Lower Level Combinations vs. Higher Level Combinations	Proportion Estimate	95% CI	
			Lower	Upper
< 13 yrs	Proportion Single Ratings $\leq$ 2 ADL Ratings	0.672	0.548	0.795
	Proportion of 2 ADL Ratings $\leq$ 3 ADL Ratings	0.858	0.777	0.939
	Proportion of 3 ADL Ratings $\leq$ 4 ADL Ratings	0.824	0.731	0.917
	Proportion of 4 ADL Ratings $\leq$ 5 ADL Ratings	0.845	0.759	0.931
	Proportion of 5 ADL Ratings $\leq$ 6 ADL Ratings	0.843	0.748	0.938
	Proportion of 6 ADL Ratings $\leq$ All 7 ADLs	0.693	0.559	0.828
	Combined			
13+ yrs	Proportion Single Ratings $\leq$ 2 ADL Ratings	0.621	0.548	0.694
	Proportion of 2 ADL Ratings $\leq$ 3 ADL Ratings	0.740	0.677	0.804
	Proportion of 3 ADL Ratings $\leq$ 4 ADL Ratings	0.736	0.673	0.800
	Proportion of 4 ADL Ratings $\leq$ 5 ADL Ratings	0.797	0.742	0.853
	Proportion of 5 ADL Ratings $\leq$ 6 ADL Ratings	0.770	0.684	0.815
	Proportion of 6 ADL Ratings $\leq$ All 7 ADLs	0.694	0.393	0.551
	Combined			